JOURNAL OF NUMBER THEORY 3, 125–149 (1971)

On Sums of Squares and on Elliptic Curves over Function Fields

J. W. S. CASSELS

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16 Mill Lane, Cambridge, England

AND

W. J. ELLISON*

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104

AND

A. PFISTER†

Mathematical Institute, University of Göttingen, Göttingen, Germany

Communicated March 11, 1970

It has long been known that every positive semidefinite function of $\mathbf{R}(x, y)$ is the sum of four squares. This paper gives the first example of such a function which is not expressible as the sum of three squares. The proof depends on the determination of the points on a certain elliptic curve defined over $\mathbf{C}(x)$. The 2-component of the Tate–Šafarevič group of this curve is nontrivial and infinitely divisible.

ЭТО НЕ ТОЛЬКО ОТРИЦАТЕЛЬНАЯ ВЕЛИЧИНА, НО ОТРИЦАТЕЛЬНАЯ ВЕЛИЧИНА ВОЗВЕДЕННАЯ В КВАДРАТ!

(ATTRIBUTED TO J. V. STALIN)

0. INTRODUCTION

Let x, y be independent indeterminates over the real field $\mathbf{R}$. Hilbert [6] showed that every positive semidefinite function of $\mathbf{R}(x, y)$ is the sum of squares of elements of $\mathbf{R}(x, y)$. Landau [16] showed that four squares suffice and it has been a long-standing problem whether or not three

* Current address: DPMMS, 16 Mill Lane, Cambridge, England.
† Current address: Mathematical Institute, University of Mainz, Mainz, Germany.

641/3/2-1

squares would suffice. The main object of this paper is to show that the positive semidefinite function

$$f(x, y) = 1 + x^2(x^2 - 3)\, y^2 + x^2 y^4, \tag{0.1}$$

exhibited by Motzkin [8] in another context, cannot be represented as a sum of three squares in $\mathbf{R}(x, y)$.

The proof is in two steps: First, we show that f is a sum of three squares if and only if a certain elliptic curve $\mathscr{C}^{-1}$ (the notation will become clear later) defined over $k = \mathbf{R}(x)$ has k-rational points with a certain additional property (Theorem 2.1). Secondly, we prove that there are no such points on $\mathscr{C}^{-1}$ by determining completely the group $\mathscr{C}_{\mathbf{C}(x)}$ of $\mathbf{C}(x)$-rational points on $\mathscr{C}$ (Theorem 7.1).

We also discuss briefly some consequences of this result for the general theory of quadratic forms and of elliptic curves over function fields.

1. Quadratic Forms in Real Function Fields

The problem of representing rational functions as sums of squares was first discussed by Hilbert [5]. Landau [16] using ideas of Hilbert [6] showed that every positive semidefinite rational function in two variables over the reals is a sum of four squares. On the other hand, it is easy to see that

$$p(x, y) = 1 + x^2 + y^2$$

is not a sum of two squares in $\mathbf{R}(x, y)$. For p is irreducible in $\mathbf{C}[x, y]$ so that any representation

$$pf_0^2 = f_1^2 + f_2^2 = (f_1 + if_2)(f_1 - if_2),$$

with $0 \neq f_0, f_1, f_2 \in \mathbf{R}[x, y]$ would imply $p \mid f_1, p \mid f_2$; which immediately leads to a contradiction.

Both results have been generalized to the n-variable case where they read as follows: Every positive semidefinite function in $\mathbf{R}(x_1, \ldots, x_n)$ is a sum of 2^n squares [1, 10]; $1 + x_1^2 + \cdots + x_n^2$ is not a sum of n squares in $\mathbf{R}(x_1, \ldots, x_n)$ [4, 2].

Let $t = t(n)$ be the smallest natural number such that every sum of squares in $\mathbf{R}(x_1, \ldots, x_n)$ is already a sum of t squares. Then $t(n)$ satisfies the inequality

$$n + 1 \leqslant t(n) \leqslant 2^n.$$

We will prove that $t(2) = 4$, but it should be pointed out that our method of proof—to translate the problem into a question about an elliptic

curve—is restricted to the case $n = 2$. We have no idea how to attack the conjecture $t(n) = 2^n$ for arbitrary n.

From now on we work in the field $\mathbf{R}(x, y)$. Let f be a positive semidefinite function in $\mathbf{R}(x, y)$. If we ask for a representation of f as a sum of squares we may clearly suppose that f is a polynomial. Since the corresponding problem for $\mathbf{R}(x)$ is trivial one may also suppose that f actually depends on x and y, i.e., cannot be written as a function of only one variable. Hence, the "easiest" cases for f to be considered are:

(a) f is a polynomial of total degree 4;

(b) f is a polynomial of degree 2 with respect to y.

In both cases f is a sum of three squares in $\mathbf{R}(x, y)$.

In case (a) this result is due to Hilbert [5], in case (b) it may be proved as follows: We can suppose that

$$f(x, y) = g(x)\, y^2 + h(x),$$

where $g = g_1^2 + g_2^2$, $h = h_1^2 + h_2^2$ in $\mathbf{R}[x]$. Put

$$f = (g_1 y + g_2 \eta)^2 + (g_2 y - g_1 \eta)^2 + \xi^2$$

with $\xi, \eta \in R(x)$. The condition on ξ, η is

$$\xi^2 + g\eta^2 = h.$$

This equation is soluble since a quadratic form of shape $(1, g)$ represents all totally positive elements of the field $\mathbf{R}(x)$ [10].

Thus, a positive semidefinite polynomial $f(x, y)$ which is not a sum of three squares must be of total degree at least 6 and of degree at least 4 in the single variables x and y. Fortunately, a promising polynomial f of this type has been discovered by Motzkin [8],[1] namely (0.1). It is the simplest known positive semidefinite polynomial which is not a sum of (any finite number of) squares of polynomials in $R[x, y]$. Other polynomials with this property have been given by R.M. Robinson [12]. There are various ways to show that $f(x, y)$ in (0.1) is positive semidefinite. Perhaps the simplest proof is to note that x^2y^2 is the geometric mean of 1, x^4y^2, and x^2y^4. Alternatively,

$$f(x, y) = \frac{(1 - x^2y^2)^2 + x^2(1 - y^2)^2 + x^2(1 - x^2)^2\, y^2}{1 + x^2}; \qquad (1.1)$$

which leads to an explicit representation of f as a sum of four squares in $\mathbf{R}(x)[y]$ on multiplying numerator and denominator with $1 + x^2$.

[1] We owe the reference to this paper to R. M. Robinson and O. Taussky-Todd.

2. A Connexion between Quadratic Forms and Elliptic Curves

Let k be a formally real field, let

$$f(y) = 1 + ay^2 + by^4 \tag{2.1}$$

$\in k[y]$ with $b \neq 0$, $a^2 \neq 4b$. Then we have:

Theorem 2.1. *$f(y)$ is a sum of three squares in $k(y)$ iff the elliptic curve*

$$\mathscr{C}^{-1} : -\eta^2 = \xi(\xi^2 - 2a\xi + a^2 - 4b) \tag{2.2}$$

has a k-rational point (ξ, η) with $\xi, \eta \in k$, such that

$$\xi \text{ and } -\xi^2 + 2a\xi - a^2 + 4b \text{ are sums of two squares in } k. \tag{2.3}$$

Proof. Suppose first that $f(y)$ is a sum of three squares in $k(y)$. Then by Ref. [2], the same is true in $k[y]$, i.e., we have

$$f = f_1^2 + f_2^2 + f_3^2 \text{ with } f_i \in k[y].$$

Since k is formally real, the f_i must be of degree $\leqslant 2$, say

$$f_i(y) = a_i + b_i y + c_i y^2 \qquad (i = 1, 2, 3).$$

Comparing coefficients we get the following system of quadratic equations in k:

$$\sum_1^3 a_i^2 = 1, \sum_1^3 a_i b_i = 0, \sum_1^3 b_i^2 + 2\sum_1^3 a_i c_i = a, \sum_1^3 b_i c_i = 0, \sum_1^3 c_i^2 = b. \tag{2.4}$$

After an orthogonal transformation over k which takes the vector (a_1, a_2, a_3) into $(1, 0, 0)$ we may assume that $a_1 = 1$, $a_2 = a_3 = 0$. Then (2.4) reduces to

$$b_1 = 0, b_2^2 + b_3^2 = a - 2c_1, b_2c_2 + b_3c_3 = 0, c_2^2 + c_3^2 = b - c_1^2. \tag{2.5}$$

This implies

$$\begin{aligned}(a - 2c_1)(b - c_1^2) &= (b_2^2 + b_3^2)(c_2^2 + c_3^2) \\ &= (b_2c_2 + b_3c_3)^2 + (b_2c_3 - b_3c_2)^2 \\ &= (b_2c_3 - b_3c_2)^2.\end{aligned} \tag{2.6}$$

Putting $\xi = a - 2c_1$, $\eta = 2(b_2c_3 - b_3c_2)$, we get $4(b - c_1^2) = 4b - (\xi - a)^2$ and

$$\xi((\xi - a)^2 - 4b) = -\eta^2. \tag{2.7}$$

From (2.5) we see that ξ and $4b - (\xi - a)^2$ are sums of two squares in k. This proves the first part of the theorem.

Conversely, if $\xi, \eta \in k$ with (2.2) and (2.3) are given, we get a solution of the system (2.5) as follows: If $\xi = 0$, $4b - a^2 = d^2 + e^2$, we take

$$b_1 = b_2 = b_3 = 0,\ 2c_1 = a,\ 2c_2 = d,\ 2c_3 = e.$$

If $\xi = b_2^2 + b_3^2 \neq 0$, we have $4b - (\xi - a)^2 = (\eta/\xi)^2 (b_2^2 + b_3^2)$. Hence we may take

$$b_1 = 0, \quad 2c_1 = a - \xi, \quad 2c_2 = \frac{\eta}{\xi} b_3, \quad 2c_3 = -\frac{\eta}{\xi} b_2.$$

Clearly every solution of (2.5) leads to a representation of $f(y)$ as a sum of three squares.

COROLLARY 2.1. *$f(x, y) = 1 + x^2(x^2 - 3)\, y^2 + x^2y^4$ is a sum of three squares in $\mathbf{R}(x, y)$ iff the elliptic curve*

$$\mathscr{C}^{-1} : -\eta^2 = \xi(\xi - x^2(x^2 - 3) - 2x)(\xi - x^2(x^2 - 3) + 2x) \tag{2.8}$$

has a rational point (ξ, η) over $\mathbf{R}(x)$ with $\eta \neq 0$ and ξ positive semidefinite.

Proof. The trivial points $(0, 0)$ and $(x^2(x^2 - 3) \pm 2x, 0)$ on (2.8) do not lead to a representation of f as a sum of three squares, since, in the first case,

$$4x^2 - (x^2(x^2 - 3))^2 = -x^2(x^2 - 1)^2 (x^2 - 4),$$

and in the two other cases

$$x^2(x^2 - 3) \pm 2x = x(x \mp 1)^2 (x \pm 2)$$

would have to be a sum of two squares in $\mathbf{R}(x)$. Furthermore, if (ξ, η) is on $\mathscr{C}^{-1}$ and $\xi \neq 0$, then (2.3) is satisfied iff ξ is a sum of two squares in $\mathbf{R}(x)$ and this is equivalent to the condition that ξ should be positive semidefinite.

COROLLARY 2.2. *The elliptic curve $\mathscr{C}^{-1}$ defined by* (2.8) *has no point (ξ, η) with $\xi, \eta \in \mathbf{Q}(x)$, $\xi \in \mathbf{Q}(x)^2$, and $\eta \neq 0$.*

Proof. The existence of such a point would imply that

$$f(x, y) = 1 + x^2(x^2 - 3)\, y^2 + x^2y^4$$

is a sum of three squares in $\mathbf{Q}(x)[y]$, say

$$f(x, y) f_0(x)^2 = f_1(x, y)^2 + f_2(x, y)^2 + f_3(x, y)^2$$

with $0 \neq f_0(x) \in \mathbf{Q}[x], f_1, f_2, f_3 \in \mathbf{Q}[x, y]$. Here we may suppose that $f_0(x)$ has no zeros in $\mathbf{Q}$ since otherwise f_0, f_1, f_2, f_3 have a common factor. Thus, one may substitute arbitrary values $u, v \in \mathbf{Q}$ for x, y and $f(u, v)$ is then a sum of three squares in $\mathbf{Q}$. But

$$f(3, 13) = 64.4159,$$

which is not a sum of three squares since $4159 \equiv -1 \bmod 8$. This concludes the proof of Corollary 2.2.

Theorem 7.1, Corollary 2.1 reduces the representability of $f(x, y)$ by the sum of three squares to a problem about elliptic curves. We shall actually do more than is required to demonstrate that no such representation exists and will determine a basis for the group of points on (2.8) defined over $\mathbf{C}(x)$. The details are given as Theorem 2 at the beginning of Section 7. Note that it is convenient to replace η by $i\eta$ in (2.8), and so to work with the curve (7.0).

3. Lemmas from the General Theory of Elliptic Curves

In this section we give some familiar results from the general theory of elliptic curves in a form suited for later application. The cognoscenti are advised to skip this section and to refer back only when necessary.

3.1. Let k and $K = k(\sqrt{d})(d \in k)$ be fields of characteristic 0 and let

$$\mathscr{C} : \eta^2 = \xi^3 + A\xi^2 + B\xi + C \tag{3.1}$$

be an elliptic curve defined over k. Let $\mathscr{C}_k$, $\mathscr{C}_K$ be the group of points on (3.1) defined over k, K, respectively, so $\mathscr{C}_k \subset \mathscr{C}_K$. Let $\mathscr{C}_k{}^d$ be the group of points of

$$\mathscr{C}^d : d\eta^2 = \xi^3 + A\xi^2 + B\xi + C$$

defined over k. Then we can regard $\mathscr{C}_k{}^d$ as a subgroup of $\mathscr{C}_K$ since $\mathscr{C}^d$ can be written

$$(\sqrt{d}\eta)^2 = \xi^2 + A\xi^2 + B\xi + C.$$

Lemma 3.1. *$2\mathscr{C}_K$ is contained in the subgroup of C_K generated by $\mathscr{C}_k$ and $\mathscr{C}_k{}^d$.*

Proof. Let σ be the automorphism $\sqrt{d} \to -\sqrt{d}$ of K/k. Then σ acts on $\mathscr{C}_K$ and $\mathfrak{a} \in \mathscr{C}_K$ is in $\mathscr{C}_k$ resp. $\mathscr{C}_k{}^d$ if

$$\sigma\mathfrak{a} = \mathfrak{a} \quad \text{resp.} \quad \sigma\mathfrak{a} = -\mathfrak{a}.$$

But now any $2\mathfrak{a} \in 2\mathscr{C}_K$ can be written

$$2\mathfrak{a} = (\mathfrak{a} + \sigma\mathfrak{a}) + (\mathfrak{a} - \sigma\mathfrak{a}).$$

3.2. Now consider the two curves

$$\mathscr{C} : \eta^2 = \xi(\xi^2 + 2A\xi + B) \tag{3.2.1}$$

and

$$\mathscr{D} : \eta_1{}^2 = \xi_1(\xi_1{}^2 + 2A_1\xi_1 + B_1), \tag{3.2.2}$$

where $A, B \in k$ and $A_1 = -2A$, $B_1 = 4A^2 - 4B$. Denote the point at infinity on $\mathscr{C}$ by $\mathfrak{o}$ and the point $(\xi, \eta) = (0, 0)$ on $\mathscr{C}$ by $\mathfrak{p}$; and similarly for $\mathfrak{o}_1$ and $\mathfrak{p}_1$ on $\mathscr{D}$. Then we have the following isogenies of degree 2:

$\phi : \mathscr{C} \to \mathscr{D}$ given by $\phi(\mathfrak{o}) = \phi(\mathfrak{p}) = \mathfrak{o}_1$ and

$$\phi(\xi, \eta) = \left(\left(\frac{\eta}{\xi}\right)^2, \frac{\xi^2 - B}{\xi^2}\,\eta\right) \quad \text{for} \quad \xi \neq 0; \tag{3.2.3}$$

$\psi : \mathscr{D} \to \mathscr{C}$ given by $\psi(\mathfrak{o}_1) = \psi(\mathfrak{p}_1) = \mathfrak{o}$ and

$$\psi(\xi_1, \eta_1) = \left(\left(\frac{\eta_1}{2\xi_1}\right)^2, \frac{\xi_1{}^2 - B_1}{8\xi_1{}^2}\,\eta_1\right) \quad \text{for} \quad \xi_1 \neq 0. \tag{3.2.4}$$

The composite maps $\psi \circ \phi$ resp. $\phi \circ \psi$ are just multiplication by 2 on $\mathscr{C}$ resp. $\mathscr{D}$.

A necessary and sufficient condition that a point $\mathfrak{a} = (a, b) \in \mathscr{C}_k$ be in $\psi(\mathscr{D}_k)$ is that

$$a \in k^2,\ a^2 + 2Aa + B \in k^2,$$

where k^2 denotes the set of squares in k. One can say rather more. Let k^* be the multiplicative group of nonzero elements of k. We have a map

$$\gamma : \mathscr{C}_k \to k^*/k^{*2} \tag{3.2.5}$$

defined as follows:

$$\begin{aligned} \gamma(\mathfrak{o}) &= 1 \cdot k^{*2}, \\ \gamma(\mathfrak{a}) &= a \cdot k^{*2}, \quad \text{if} \quad a \neq 0, \\ &= (a^2 + 2Aa + B)\,k^{*2} \quad \text{if} \quad a^2 + 2Aa + B \neq 0. \end{aligned}$$

The two definitions on the right-hand side are both applicable if $b \neq 0$ and then they coincide, and one of the two definitions is always applicable for $\mathfrak{a} \neq \mathfrak{o}$.

LEMMA 3.2. *γ is a group homomorphism. Its kernel is precisely $\psi(\mathscr{D}_k)$.*

Proof. See Ref. [14].

Replacing $\mathscr{C}$ by $\mathscr{D}$ we have a map $\delta : \mathscr{D}_k \to k^*/k^{*2}$ for which the analog of Lemma 3.2 holds.

3.3. Now suppose that $\mathscr{C}$ has the shape

$$\mathscr{C} : \eta^2 = (\xi - e_1)(\xi - e_2)(\xi - e_3), \tag{3.3.1}$$

where

$$e_1, e_2, e_3 \in k.$$

To $(a, b) \in \mathscr{C}_k$ with $b \neq 0$ there correspond three $a_j \in k^*/(k^*)^2$ given by

$$\alpha_j = (a - e_j) \circ k^{*2}. \tag{3.3.2}$$

Further

$$\alpha_1\alpha_2\alpha_3 = 1 \circ k^{*2} \tag{3.3.3}$$

by (3.3.1). When $b = 0$, only two of the α_j in (3.3.2) are well defined and the third is defined by (3.3.3). We thus have an everywhere-defined map γ_2 of $\mathscr{C}_k$ into three copies of $k^*/(k^*)^2$.

LEMMA 3.3. *γ_2 is a group homomorphism. Its kernel is precisely $2\mathscr{C}_k$.*

3.4. We are particularly concerned with a groundfield $k = k_0(x)$, where x is transcendental. The curve

$$\mathscr{C} : \eta^2 = \xi^2 + A\xi^2 + B\xi + C, \qquad A, B, C \in k, \tag{3.4.1}$$

is birationally equivalent to one defined over k_0 iff there is a linear transformation

$$\eta = l\eta_1, \quad \xi = m\xi_1 + n \qquad (l, m, n \in k)$$

such that $\eta_1^2 \in k_0(\xi_1)$. [See, e.g., 3, p. 212.] The analog of the Mordell–Weil finite basis theorem [7] implies the following:

LEMMA 3.4. *Suppose that $k = k_0(x)$ and that $\mathscr{C}$ is not birationally equivalent to a curve defined over k_0. Then $\mathscr{C}_k$ is finitely generated.*

3.5. In $k_0(x)$ the ring $k_0[x]$ of polynomials is an analog of the integers in the classical theory over the rationals. The analog of theorems by Nagell and Lutz [3] is

LEMMA 3.5. *In* (3.4.1), *suppose that*

$$A, B, C \in k_0[x]$$

and that $\mathfrak{a} = (a, b) \in \mathscr{C}_k$ *is of finite order. Then*

$$a, b \in k_0[x].$$

Further, either $b = 0$ *or* b^2 *divides the discriminant* $\Delta(\mathscr{C})$ of

$$\xi^3 + A\xi^2 + B\xi + C.$$

4. POINTS OF FINITE ORDER

4.1. Let now $A = -x^2(x^2 - 3)$, $B = x^2(x^2 - 1)^2 (x^2 - 4)$, $A_1 = 2x^2(x^2 - 3)$, and $B_1 = 16x^2$ in Section 3.2, i.e.,

$$\begin{aligned} \mathscr{C}\colon \eta^2 &= \xi(\xi^2 - 2x^2(x^2 - 3)\,\xi + x^2(x^2 - 1)^2 (x^2 - 4)) \\ &= \xi(\xi - x^2(x^2 - 3) - 2x)(\xi - x^2(x^2 - 3) + 2x), \end{aligned} \tag{4.1}$$

$$\mathscr{D}\colon \eta_1^2 = \xi_1(\xi_1^2 + 4x^2(x^2 - 3)\,\xi_1 + 16x^2). \tag{4.2}$$

We intend to determine all points of finite order on $\mathscr{C}$ and $\mathscr{D}$ defined over $k = \mathbf{C}(x)$. There are clearly the following points of order 2:

$\mathfrak{p} = (0, 0)$, $\mathfrak{q} = (x^2(x^2 - 3) + 2x, 0)$, $\mathfrak{r} = (x^2(x^2 - 3) - 2x, 0)$ on $\mathscr{C}$, $\mathfrak{p}_1 = (0, 0)$ on $\mathscr{D}$.

[The two other points of order 2 on $\mathscr{D}$ which are given by

$$\xi_1^2 + 4x^2(x^2 - 3)\,\xi_1 + 16x^2 = 0$$

are not rational over $\mathbf{C}(x)$.] The image under γ of $\mathfrak{p}$, $\mathfrak{q}$, and $\mathfrak{r}$ is not a square in k, hence $\mathfrak{p}, \mathfrak{q}, \mathfrak{r} \notin \psi(\mathscr{D}_k)$ by Lemma 3.2. *A fortiori* there are no points of order 4 on $\mathscr{C}_k$ and $\mathscr{D}_k$. (Use $\mathfrak{p}_1 = \phi(\mathfrak{q}) = \phi(\mathfrak{r})$.)

It remains to determine the points of odd order. Suppose that $\mathfrak{a}_1 = (a_1, b_1) \in \mathscr{D}_k$ is of odd order $n > 1$. Then $\mathfrak{a} = \psi(\mathfrak{a}_1) = (a, b) \in \mathscr{C}_k$ is also of order n. By Lemma 3.5 we have

$$a, b, a_1, b_1 \in \mathbf{C}[x],\ b, b_1 \neq 0 \text{ and } b^2 \mid \Delta(\mathscr{C}),\ b_1^2 \mid \Delta(\mathscr{D}).$$

In addition, $a_1 \in \mathbf{C}[x]^2$ since $\mathfrak{a}_1 = 2((n+1)/2\,\mathfrak{a}_1) \in \phi(\mathscr{C}_k)$. By (3.2.4),

$$a = \frac{{b_1}^2}{4{a_1}^2} = \frac{{a_1}^2 + 4x^2(x^2-3)\,a_1 + 16x^2}{4a_1},$$

hence $a_1 \mid x^2$. Further, the point $\mathfrak{a}_1 + \mathfrak{p}_1 = \mathfrak{a}_1' = (a_1', b_1') \in \mathscr{D}_k$ has order $2n$ and an easy calculation shows that $a_1' = 16x^2/a_1$. Replacing $\mathfrak{a}_1$ by $\mathfrak{a}_1'$, if necessary, we may, therefore, assume that $a_1 \in C^*$. Then

$${b_1}^2 = 4{a_1}^2x^4 + 4a_1(4 - 3a_1)\,x^2 + {a_1}^3$$

and

$${b_1}^2 \mid \Delta(\mathscr{D}) = 4B_1^2(B_1 - A_1^2) = -16B_1^2B = -16^3x^6(x^2-1)^2\,(x^2-4).$$

This implies

$$\begin{aligned} {b_1}^2 &= 4{a_1}^2(x^2-1)^2, \\ a_1 &= 4, \qquad b_1 = \pm 8(x^2-1). \end{aligned}$$

Put

$$\mathfrak{s}_1 = (4, 8(x^2-1)) \in \mathscr{C}_k\,. \tag{4.3}$$

By Lemma 3.2, $\mathfrak{s}_1 \in \phi(\mathscr{C}_k)$ and an easy calculation shows that the pre-images of $\mathfrak{s}_1$ under ϕ are

$$\mathfrak{s} = (x^2(x^2-1),\, 2x^2(x^2-1)) \in \mathscr{C}_k \tag{4.4}$$

and $\mathfrak{s} + \mathfrak{p}$. We also note that

$$2\mathfrak{s} = \psi(\mathfrak{s}_1) = ((x^2-1)^2, -(x^2-1)^2), \tag{4.5}$$

$$2\mathfrak{s}_1 = \phi(2\mathfrak{s}) = (1, -(2x^2+1)). \tag{4.6}$$

By assumption $\mathfrak{s}_1$ has order n or $2n$, hence $2\mathfrak{s}_1$ has order n. But this is impossible since

$$2x^2 + 1 \nmid \Delta(\mathscr{D}).$$

Therefore $\mathfrak{s} \in \mathscr{C}_k$ and $\mathfrak{s}_1 \in \mathscr{D}_k$ are points of infinite order. We have proved:

LEMMA 4.1. *Let $k = \mathbf{C}(x)$ and denote the torsion subgroup of C_k* resp. *$\mathscr{D}_k$ by $\mathscr{C}_{k,0}$* resp. *$\mathscr{D}_{k,0}$. Then $\mathscr{C}_{k,0} \cong \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, with generators $\mathfrak{p}, \mathfrak{q}$, $\mathscr{D}_{k,0} \cong \mathbf{Z}/2\mathbf{Z}$, with generator $\mathfrak{p}_1$.*

4.2. For later reference we will also prove that the points $\mathfrak{s}_1$, $\mathfrak{s}_1 + \mathfrak{p}_1 \in \mathscr{D}_k$ are not divisible in $\mathscr{D}_k$ by any number $n > 1$. This is clear for $n = 2$, since $\mathfrak{s}$, $\mathfrak{s} + \mathfrak{p}$, $\mathfrak{s} + \mathfrak{q}$, and $\mathfrak{s} + \mathfrak{r}$ are not in $\psi(\mathscr{D}_k)$ by

Lemma 3.2. So let n be an odd integer >1. It is enough to show that $2\mathfrak{s}_1 = (1, -(2x^2 + 1))$ is not divisible by n in $\mathscr{D}_k$. Suppose

$$2\mathfrak{s}_1 = n\mathfrak{a}_1 , \quad \mathfrak{a}_1 \in \mathscr{D}_k . \tag{4.7}$$

The group of automorphisms σ of $\mathbf{C}$ over $\mathbf{Q}$ acts in a natural way on $\mathscr{D}_k$ and leaves $\mathfrak{s}_1$ invariant since $\mathfrak{s}_1$ is defined over $\mathbf{Q}(x)$. Hence

$$n(\sigma\mathfrak{a}_1 - \mathfrak{a}_1) = 0.$$

But $\mathscr{D}_k$ contains no points of odd order, so $\sigma\mathfrak{a}_1 = \mathfrak{a}_1$, i.e., $\mathfrak{a}_1$ is defined over $\mathbf{Q}(x)$.

From the general theory of elliptic curves ([3], Lemma 7.2) we have the following results: Let $\mathfrak{x}_1 = (\xi_1 , \eta_1)$ be a generic point on $\mathscr{D}$:

$$\eta_1{}^2 = \xi_1(\xi_1{}^2 + 2A_1\xi_1 + B_1).$$

Then $2\mathfrak{x}_1 = \mathfrak{x}_2 = (\xi_2 , \eta_2)$ with

$$\xi_2 = \left(\frac{\xi_1{}^2 - B_1}{2\eta_1}\right)^2 = \frac{\phi_2(\xi_1)}{4\eta_1{}^2} , \qquad \eta_2 = \frac{\Omega_2(\xi_1)}{8\eta_1{}^3} ; \tag{4.8}$$

$n\mathfrak{x}_1 = \mathfrak{x}_n = (\xi_n , \eta_n)$ with

$$\xi_n = \frac{\phi_n(\xi_1)}{\psi_n{}^2(\xi_1)} , \qquad \eta_n = \eta_1 \frac{\Omega_n(\xi_1)}{\psi_n{}^3(\xi_1)} , \tag{4.9}$$

where ϕ_n, ψ_n, and Ω_n are polynomials with coefficients in $\mathbf{Q}[A_1 , B_1]$, ϕ_n of degree n^2 with top coefficient 1, ψ_n of degree $\frac{1}{2}(n^2 - 1)$ with top coefficient n.

From the addition formula we have

$$\begin{aligned}\xi_{n+2} &= \frac{\xi_2\xi_n(\xi_2 + \xi_n + 4A_1) + B_1(\xi_2 + \xi_n) - 2\eta_2\eta_n}{(\xi_n - \xi_2)^2} \\ &= \frac{\left\{\begin{array}{r}\phi_2\phi_n(\phi_2\psi_n{}^2 + \phi_n \cdot 4\eta_1{}^2 + 4A_1\psi_n{}^2 \cdot 4\eta_1{}^2) \\ + B_1 \cdot 4\eta_1{}^2\psi_n{}^2(\phi_2\psi_n{}^2 + \phi_n \cdot 4\eta_1{}^2)\end{array}\right\}}{(\phi_n \cdot 4\eta_1{}^2 - \phi_2 \cdot \psi_n{}^2)^2} \\ &\quad - \frac{4\eta_1{}^2\Omega_2\Omega_n\psi_n}{(\phi_n \cdot 4\eta_1{}^2 - \phi_2\psi_n{}^2)^2} .\end{aligned} \tag{4.10}$$

This shows that

$$\psi_{n+2}(\xi_1)| \, 4\xi_1(\xi_1{}^2 + 2A_1\xi_1 + B_1)\, \phi_n(\xi_1) - (\xi_1{}^2 - B_1)^2\, \psi_n{}^2(\xi_1). \tag{4.11}$$

We can now prove by a specialization argument that $\phi_n(0) = 0$ and that $\psi_n(0)$ is a power (depending on n) of B_1. For, if $\mathfrak{x}_1 = \mathfrak{p}_1 = (0, 0)$ we must have $\mathfrak{x}_n = n\mathfrak{p}_1 = \mathfrak{p}_1$, hence $\xi_n = 0$ whenever $\xi_1 = 0$, i.e.,

$$\phi_n(0) = 0. \tag{4.12}$$

From (4.11)

$$\psi_{n+2}(0) \mid B_1^2 \psi_n^2(0)$$

and since $\psi_1(0) = 1$ and $B_1 = 16x^2$ this shows by induction that $\psi_n(0)$ is a power of x (up to a constant factor) for any odd number n.

In our assumption (4.7), let

$$\mathfrak{a}_1 = \left(\frac{a_1}{c_1^2}, \frac{b_1}{c_1^3}\right)$$

with $a_1, b_1, c_1 \in \mathbf{Q}[x]$, and g.c.d. $(a_1, c_1) = 1$. Then

$$1 = \frac{\phi_n(a_1/c_1^2)}{\psi_n^2(a_1/c_1^2)} = \frac{a_1^{n^2} + \cdots + \phi_n(0)\, c_1^{2n^2}}{(na_1^{1/2(n^2-1)} + \cdots + \psi_n(0)\, c_1^{n^2-1})^2 \cdot c_1^2}.$$

This implies $c_1 = 1$ and $a_1 \mid \psi_n(0)^2$, i.e., a_1 divides a power of $B_1 = 16x^2$. From (4.7) and Lemma (3.2) $a_1 \in \mathbf{Q}[x]^2$. Suppose now that

$$a_1 = a^2x^{2k}, \quad a \in \mathbf{Q}^*, \quad k \geqslant 0. \tag{4.13}$$

If $k \geqslant 1$,

$$a_1^2 + 4x^2(x^2 - 3)\, a_1 + 16x^2 = x^2(a^4x^{4k-2} + 4a^2(x^2 - 3)\, x^{2k} + 16)$$

has to be a square, say

$$= x^2(4 - \tfrac{3}{2}a^2x^{2k} + \text{terms of degree } > 2k \text{ in } x)^2.$$

This is clearly impossible for $k > 1$. For $k = 1$, we find

$$\begin{aligned} a^4x^2 + 4a^2x^4 - 12a^2x^2 + 16 &= (2ax^2 \pm 4)^2, \\ a^4 - 12a^2 &= \pm 16a, \\ a(a^2 - 12) &= \pm 16, \\ a &= \pm 2, \pm 4, \\ a_1 &= 4x^2 \text{ or } a_1 = 16x^2. \end{aligned}$$

If $k = 0$, we find

$$\begin{aligned} a^4 + 4a^2x^4 - 12a^2x^2 + 16x^2 &= (a^2 \pm 2ax^2)^2, \\ -12a^2 + 16 &= \pm 4a^3, \\ a^2(3 \pm a) &= 4, \\ a &= \pm 1, \pm 2, \\ a_1 &= 1 \text{ or } a_1 = 4. \end{aligned}$$

The solutions $a_1 = 1, 4, 4x^2, 16x^2$ correspond to the points

$$\mathfrak{a}_1 = \pm 2\mathfrak{s}_1\,,\ \pm\mathfrak{s}_1\,,\ \pm\mathfrak{s}_1 + \mathfrak{p}_1\,,\ \pm\ 2\mathfrak{s}_1 + \mathfrak{p}_1\,,$$

respectively; which do not satisfy the Eq. (4.7). We have

LEMMA 4.2. *Let* $k = \mathbf{C}(x)$. *The points* $\mathfrak{s}_1$ *and* $\mathfrak{s}_1 + \mathfrak{p}_1$ *are not divisible in* $\mathscr{D}_k$. *The points* $\mathfrak{s}$, $\mathfrak{s} + \mathfrak{p}$, $\mathfrak{s} + \mathfrak{q}$, *and* $\mathfrak{s} + \mathfrak{r}$ *are not divisible in* $\mathscr{C}_k$.

5. FIRST DESCENTS OVER $\mathbf{R}(x)$

5.0. In this section we look at the first descents for $\mathscr{C}$ and $\mathscr{C}^{-1}$. It is convenient to decompose multiplication by 2 into the product of two isogenies along the lines of Section 3.2 by considering the curve

$$\mathscr{D} : \eta^2 = \xi(\xi^2 + 4x^2(x^2 - 3)\,\xi + 16x^2) \tag{5.0.1}$$

as well as $\mathscr{C}$.

It is convenient to work in $\mathbf{R}[x]$ rather than in $\mathbf{R}(x)$. By abuse of language we denote by (ξ, η) a point of the curve we are considering defined over $\mathbf{R}(x)$. On writing ξ as a quotient of polynomials, one readily sees that the denominator must be a perfect square, say

$$\xi = \omega/\psi^2, \tag{5.0.2}$$

where

$$\omega, \psi \in \mathbf{R}[x],\ \text{g.c.d.}\ (\omega, \psi) = 1. \tag{5.0.3}$$

Considerations of divisibility show that

$$\omega = f\theta^2, \tag{5.0.4}$$

where

$$\theta \in \mathbf{R}[x],\ \text{g.c.d.}(f, \psi) = \text{g.c.d.}(\theta, \psi) = 1 \tag{5.0.5}$$

and

$$f \in R[x] \tag{5.0.6}$$

is one of a finite set. We endeavour to make this finite set as small as possible. By Lemma 3.2, a bound on the number of f gives a bound on the cokernel of the corresponding isogeny.

5.1. On applying the above substitution to ξ in the equation for $\mathscr{C}$ we obtain

$$\omega\{\omega^2 - 2x^2(x^2 - 3)\,\omega\psi^2 + x^2(x^2 - 1)^2\,(x^2 - 4)\,\psi^4\} = \text{square}. \tag{5.1.1}$$

The greatest common divisor of the two factors on the left side divides $x^2(x^2-1)^2(x^2-4)$ and so

$$\omega = f\theta^2, \quad f \mid x(x+1)(x-1)(x+2)(x-2) \tag{5.1.2}$$

and

$$f^2\theta^4 - 2x^2(x^2-3)f\theta^2\psi^2 + x^2(x^2-1)^2(x^2-4)\psi^4 = +f\lambda^2 \ (\lambda \in R[x]). \tag{5.1.3}$$

Suppose first that the coefficient of the highest power of x in f is negative. Then the coefficient of the highest power of x on the left side of (5.1.3) is positive, while that on the right side is negative; a contradiction. Hence the highest coefficient is positive.

Now suppose that f has odd degree, so the left side of (5.1.3) has odd degree (because the right has). Then the degrees of the three summands on the left side of (5.1.3) are unequal and either the first or the third has the highest degree; a contradiction.

Thus, f is of even degree with positive highest coefficient. By (5.1.2), the group generated by the f in $(\mathbf{R}(x))^*/((\mathbf{R}(x))^*)^2$ has thus at most four generators, of which we can account for three:

LEMMA 5.1. *Independent generators of*

$$\operatorname{coker}(\mathscr{D}_{\mathbf{R}(x)} \to \mathscr{C}_{\mathbf{R}(x)}) \tag{5.1.4}$$

are given by the point

$$\mathfrak{s} = (x^2(x^2-1), 2x^2(x^2-1)) \tag{5.1.5}$$

and the two points

$$\begin{aligned} \mathfrak{q} &= (x(x-1)^2(x+2), 0), \\ \mathfrak{r} &= (x(x+1)^2(x-2), 0) \end{aligned} \tag{5.1.6}$$

of order 2. There is at most one further generator, corresponding to $f = x(x-1)$.

Lemma 3.2 ensures, of course, that the three specified points give independent generators of the cokernel.

5.2. We now consider (5.0.1) similarly and make the substitution (5.0.2). Then

$$\omega(\omega^2 + 4x^2(x^2-3)\omega\psi^2 + 16x^2\psi^4) = \text{square}. \tag{5.2.1}$$

The greatest common divisor of the factors on the r.h.s. divides x^2 and so

$$\omega = f\theta^2, f \mid x \tag{5.2.2}$$

and

$$f^2\theta^4 + 4x^2(x^2 - 3)f\theta^2\psi^2 + 16x^2\psi^4 = +f\lambda^2 \tag{5.2.3}$$

with

$$\lambda \in R[x]. \tag{5.2.4}$$

The possibilities $f = -1$ and $f = \pm x$ lead to a contradiction, on considering the sign of the lowest power of x on both sides of (5.2.3). Hence,

LEMMA 5.2. $\operatorname{coker}(\mathscr{C}_{\mathbf{R}(x)} \to \mathscr{D}_{\mathbf{R}(x)})$ *is trivial.*

Since our curve $\mathscr{C}$ is not birationally equivalent to one defined over $\mathbf{C}$, the following lemma follows on combining Lemmas 3.4, 4.1, 5.1, and 5.2.

LEMMA 5.3. *The group* $\mathscr{C}_{\mathbf{R}(x)}$ *is the direct sum of two cycles of order* 2 *and either two or one cycle of infinite order according as there is or is not a point* $(\xi, \eta) \in \mathscr{C}_{\mathbf{R}(x)}$ *with* $\xi \in x(x-1)\,\mathbf{R}(x)^{*^2}$.

5.3. We now consider the same problem for $\mathscr{C}^{-1}$ in which $-\eta^2$ is substituted for η^2. As in Section 5.1, we have to consider

$$f^2\theta^4 + 2x^2(x^2 - 3)f\theta^2\psi^2 + x^2(x^2 - 1)^2(x^2 - 4)\psi^4 = +f\lambda^2, \tag{5.3.1}$$

where

$$\xi = \frac{\omega}{\psi^2}, \qquad \omega = -f\theta^2,$$

and

$$f \mid x(x+1)(x-1)(x+2)(x-2). \tag{5.3.2}$$

As in Section 5.1, the degree of f is even.

We now show that $(x-1) \mid f$ is impossible. Suppose

$$f = (x-1)\,g,\ g \mid x(x+1)(x+2)(x-2).$$

Dividing by $x - 1$ and then putting $x = 1$, we get from (5.3.1)

$$-4g(1)\,\theta(1)^2\,\psi(1)^2 = g(1)\,\lambda(1)^2. \tag{5.3.3}$$

Since g.c.d. $(f, \psi) = 1$ we have $\psi(1) \neq 0$, hence $\theta(1) = \lambda(1) = 0$. But then the r.h.s of (5.3.1) is divisible by $(x-1)^3$ and the l.h.s is not.

Similarly, $(x+1) \mid f$ is impossible. Thus, the degree of f is even and $f \mid x(x+2)(x-2)$. As in Section 5.1 the points with $\eta = 0$ give the possibilities

$$f = x(x+2),\ x(x-2).$$

If any other value of f occurs, then so must $f = -1$, since the possible values of f are a group modulo $R(x)^{*2}$. We have thus proved the following lemma :

LEMMA 5.4. $\operatorname{coker}(\mathscr{D}^{-1}_{\mathbf{R}(x)} \to \mathscr{C}^{-1}_{\mathbf{R}(x)})$ *has three or two generators according as there is or is not a point of* $\mathscr{C}^{-1}_{\mathbf{R}(x)}$ *with* $\xi \in \mathbf{R}(x)^{*2}$. *Two generators are given by the points* (5.1.6) *of order* 2.

The same arguments as in Section 5.2 give

LEMMA 5.5. $\operatorname{coker}(\mathscr{C}^{-1}_{\mathbf{R}(x)} \to \mathscr{D}^{-1}_{\mathbf{R}(x)})$ *is trivial.*

Combining Lemmas 5.4 and 5.5 with Lemma 3.4 and 4.1, we have the following:

LEMMA 5.6. $\mathscr{C}^{-1}_{\mathbf{R}(x)}$ *is the direct sum of two groups of order* 2 *and either one or no cyclic group of infinite order according as there is or is not a point* $(\xi, \eta) \in \mathscr{C}^{-1}_{\mathbf{R}(x)}$ *with* $\xi \in R(x)^{*2}$.

In the latter case where there is no such point (ξ, η) we have clearly $\mathscr{C}^{-1}_{\mathbf{R}(x)} \subset \mathscr{C}_{\mathbf{R}(x)}$. By Lemma 3.1 this implies $2\mathfrak{a} \in \mathscr{C}_{\mathbf{R}(x)}$ for any $\mathfrak{a} \in \mathscr{C}_{\mathbf{C}(x)}$ or, if ρ denotes the automorphism "complex conjugation", $2(\mathfrak{a} - \rho\mathfrak{a}) = \mathfrak{o}$. Thus $\mathfrak{a} - \rho\mathfrak{a}$ is one of the four 2-division points. On the other hand, $\mathfrak{a}$ and $\rho\mathfrak{a}$ have the same image under γ in $\mathbf{C}(x)^*/\mathbf{C}(x)^{*2}$, since $\gamma(\mathscr{C}_{\mathbf{C}(x)})$ is generated by square-classes which are defined over $\mathbf{Q}(x)$ as we see from (5.1.2). So we must have $\mathfrak{a} = \rho\mathfrak{a}$, i.e., $\mathfrak{a} \in \mathscr{C}_{\mathbf{R}(x)}$.

Finally, we have proved the following:

LEMMA 5.7. $\mathscr{C}_{\mathbf{C}(x)}$ *is the direct sum of two groups of order* 2 *and either* 1, 2 *or* 3 *infinite cyclic groups. Further,* $\mathscr{C}_{\mathbf{C}(x)} = \mathscr{C}_{\mathbf{R}(x)}$ *unless there is a point* (ξ, η) *of* $\mathscr{C}^{-1}_{\mathbf{R}(x)}$ *with* $\xi \in \mathbf{R}(x)^{*2}$.

It is perhaps worth remarking that we could have obtained the first sentence of Lemma 5.7 more easily by doing the descents in $\mathbf{C}(x)$; which provides a check on the preceding argument.

6. THE ACTION OF GALOIS

The curve $\mathscr{C}$ itself is defined over $\mathbf{Q}(x)$. The elements of $\mathscr{C}_{\mathbf{C}(x)}$ are each defined over a finite extension of $\mathbf{Q}(x)$ since otherwise one could get uncountably many elements of $\mathscr{C}_{\mathbf{C}(x)}$ by specialization, contrary to

Lemma 3.4. Let K be the smallest extension of $\mathbf{Q}$ such that a set of generators of $\mathscr{C}_{\mathbf{C}(x)}$ is defined over $K(x)$. Then K is a finite extension and every element of $\mathscr{C}_{\mathbf{C}(x)}$ is defined over $K(x)$. Further, $K/\mathbf{Q}$ is normal since the conjugate of a point of $\mathscr{C}_{\mathbf{C}(x)}$ is also in $\mathscr{C}_{\mathbf{C}(x)}$. The Galois group Γ (say) of $K/\mathbf{Q}$ clearly acts faithfully on $\mathscr{C}_{\mathbf{C}(x)}$.

By Lemmas 4.1, 4.2, and 5.7 a set of generators of $\mathscr{C}_{\mathbf{C}(x)}$ can be chosen such as to contain the points $\mathfrak{p}$, $\mathfrak{q}$ (generators for $\mathscr{C}_{\mathbf{C}(x),0}$), and $\mathfrak{s}$. Let

$$\mathfrak{H} \subset \mathbf{C}(x) \tag{6.1}$$

denote the subgroup of $\mathscr{C}_{\mathbf{C}(x)}$ which is generated by $\mathfrak{p}$, $\mathfrak{q}$, and $\mathfrak{s}$. Then the factor group

$$\mathfrak{F} = \mathscr{C}_{\mathbf{C}(x)}/\mathfrak{H}. \tag{6.2}$$

is a direct product of 0, 1 or 2 infinite cyclic groups and Γ acts on $\mathfrak{F}$, since the action of Γ on $\mathfrak{H}$ is trivial.

Further information is given by the following two lemmas.

LEMMA 6.1. *Γ acts trivially on $\mathscr{C}_{\mathbf{C}(x)}/2\,\mathscr{C}_{\mathbf{C}(x)}$.*

For it is easy to verify that Γ acts trivially on the image of γ_2 in Lemma 3.3, when the $\mathscr{C}$ of (3.3.1) is identified with the present $\mathscr{C}$ and $k = \mathbf{C}(x)$. Indeed by a consideration of factorization like that of Section 5, one can verify that representatives of the relevant classes of $\mathbf{C}(x)^*/\mathbf{C}(x)^{*2}$ can be chosen in $\mathbf{Q}(x)$.

LEMMA 6.2. *The action of Γ on $\mathfrak{F}$ is faithful. The induced action on $\mathfrak{F}/2\mathfrak{F}$ is trivial.*

For suppose that $\sigma \in \Gamma$ acts trivially on $\mathfrak{F}$. Let $\mathfrak{c}$ be any element of $\mathscr{C}_{\mathbf{C}(x)}$. Then

$$\sigma\mathfrak{c} = \mathfrak{c} + \mathfrak{h}, \tag{6.3}$$

with $\mathfrak{h} \in \mathfrak{H}$ and $\mathfrak{h} \in 2\mathscr{C}_{\mathbf{C}(x)}$ by Lemma 6.1. Hence, by Lemma 4.2,

$$\mathfrak{h} = 2\mathfrak{k} \tag{6.4}$$

for some $\mathfrak{k} \in \mathfrak{H}$. Let n be the order of σ. Then

$$\mathfrak{c} = \sigma^n\mathfrak{c} = \mathfrak{c} + 2n\mathfrak{k}. \tag{6.5}$$

Thus $2n\mathfrak{k} = \mathfrak{o}$ and so $2\mathfrak{k} = \mathfrak{o}$ by the properties of $\mathfrak{H}$. Hence σ acts trivially on $\mathscr{C}_{\mathbf{C}(x)}$ and $\sigma = 1$.

The second sentence of the enunciation follows at once from Lemma 6.1.

We now consider the abstract situation revealed by Lemma 6.2.

LEMMA 6.3. *Let Γ be any finite group and $\mathfrak{F}$ a torsionfree module of rank at most 2 on which Γ acts faithfully. Suppose that the induced action on $\mathfrak{F}/2\mathfrak{F}$ is trivial. Then Γ is either trivial, of order 2, or noncyclic of order 4.*

The proof, when $\mathfrak{F}$ has rank 0 or 1, is simpler than when rank $\mathfrak{F} = 2$, which we shall suppose from now on. Let f_1 and f_2 be a basis and suppose that $\sigma \in \Gamma$. Then

$$\sigma f_1 = (1 + 2a) f_1 + 2bf_2 ,$$
$$\sigma f_2 = 2cf_1 + (1 + 2d) f_2 ,$$

where

$$a, b, c, d \in \mathbf{Z}.$$

The matrix

$$M = \begin{pmatrix} 1 + 2a & 2b \\ 2c & 1 + 2d \end{pmatrix}$$

has finite order, and so either

$$M = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

or the sum of the eigenvalues is less than 2 in absolute value, i.e.,

$$|(1 + 2a) + (1 + 2d)| < 2.$$

Hence

$$(1 + 2a) + (1 + 2d) = 0$$

and so

$$\det(M) = (1 + 2a)(1 + 2d) - 4cd$$
$$\equiv -1 \pmod 4,$$

that is

$$\det(M) = -1.$$

Hence, the eigenvalues of σ and M are $+1$ and -1. Let $g_+ , g_- \in \mathfrak{F}$ be bases of the one-dimensional submodules belonging to the eigenvalues $+1$, -1, respectively. We want to show that g_+ , g_- is a basis for $\mathfrak{F}$. For any $f \in \mathfrak{F}$ we have

$$2f = (f + \sigma f) + (f - \sigma f)$$

and so

$$2f = ag_+ + bg_- , \qquad a, b \in \mathbf{Z}.$$

Since $\mathfrak{F}$ is torsion-free, it is enough to show that a and b are both even.

Suppose, first that $2 \neq a$ and $2|b$. Then there is an $f_3 \in \mathfrak{F}$ with

$$2f_3 = g_+ .$$

Then

$$2\sigma f_3 = 2f_3$$

so

$$\sigma f_3 = f_3,$$

in contradiction to the definition of g_+. Similarly, we cannot have $2 \mid a$ $2 \nmid b$. Finally, if $2 \nmid a$ and $2 \nmid b$, then there is an $f_4 \in \mathfrak{F}$ with

$$2f_4 = g_+ + g_-$$

and so

$$2\sigma f_4 = g_+ - g_- = 2f_4 - 2g_-,$$

$$\sigma f_4 = f_4 - g_-,$$

in contradiction to the assumption that σ is trivial on $\mathfrak{F}/2\mathfrak{F}$. By suitable choice of base we can thus ensure that

$$M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Now let τ be another element of Γ, say corresponding to

$$M' = \begin{pmatrix} 1 + 2a', & 2b' \\ 2c' & 1 + 2d' \end{pmatrix} \neq \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and $\neq \pm M$.

Then, as before,

$$(1 + 2a') + (1 + 2d') = 0.$$

On the other hand, $\sigma\tau$ corresponds to the matrix

$$MM' = \begin{pmatrix} 1 + 2a' & 2b' \\ -2c' & -(1 + 2d') \end{pmatrix}.$$

Hence

$$(1 + 2a') - (1 + 2d') = 0.$$

Thus

$$1 + 2a' = 1 + 2d' = 0;$$

a contradiction.

This concludes the proof of the Lemma.

Lemmas 6.2 and 6.3 together give

LEMMA 6.4. *$\mathscr{C}_{\mathbf{C}(x)} = \mathscr{C}_{K(x)}$ for some normal extension K of $\mathbf{Q}$ whose Galois group is either trivial, of order 2, or noncyclic of order 4.*

Finally, we have the following refined version of Lemma 5.7:

LEMMA 6.5. *$\mathscr{C}_{\mathbf{C}(x)}$ is generated by the points $\mathfrak{p}, \mathfrak{q}$, and $\mathfrak{s}$ defined over $\mathbf{Q}(x)$ and at most two further points $\mathfrak{c}_1$ and $\mathfrak{c}_2$ defined over $K(x)$. The elements of the Galois group Γ of $K/\mathbf{Q}$ act like ± 1 on $\mathfrak{c}_1$ and $\mathfrak{c}_2$.*

For the proof we may suppose that $\mathfrak{F}$ has rank 2, the other cases being simpler. By the proof of Lemma 6.3, $\mathfrak{F}$ has a basis f_1, f_2 such that the elements of Γ operate like $\pm\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$ or $\pm\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$. Let $\mathfrak{c}_i \in \mathscr{C}_{\mathbf{C}(x)}$ be representatives for f_i $(i = 1, 2)$. Suppose also that Γ is nontrivial and that $\rho \in \Gamma$ satisfies

$$\rho f_1 = -f_1 .$$

Let $\Gamma_0 \subset \Gamma$ be the subgroup consisting of the $\sigma \in \Gamma$ such that

$$\sigma f_1 = f_1 .$$

Then Γ_0 is of index 2 in Γ. By Lemma 6.2,

$$\rho\mathfrak{c}_1 = -\mathfrak{c}_1 + 2\mathfrak{h}$$

for some $\mathfrak{h} \in \mathfrak{H}$. On replacing $\mathfrak{c}_1$ by $\mathfrak{c}_1 - \mathfrak{h}$ and recalling that Γ acts trivially on $\mathfrak{H}$, we have

$$\rho\mathfrak{c}_1 = -\mathfrak{c}_1 .$$

The argument used to prove Lemma 6.2 shows that

$$\sigma\mathfrak{c}_1 = \mathfrak{c}_1$$

for all $\sigma \in \Gamma_0$. Thus $\tau\mathfrak{c}_1 = \pm\mathfrak{c}_1$ for all $\tau \in \Gamma$. Similarly, $\tau\mathfrak{c}_2 = \pm\mathfrak{c}_2$ for all τ if $\mathfrak{c}_2$ is a suitable representative for f_2.

7. PROOF OF THEOREM 7.1

We are now able to determine the group $\mathscr{C}_{\mathbf{C}(x)}$ of $\mathbf{C}(x)$-rational points on our curve

$$\mathscr{C} : \eta^2 = \xi(\xi - x^2(x^2 - 3) - 2x)(\xi - x^2(x^2 - 3) + 2x). \qquad (7.0)$$

THEOREM 7.1. *$\mathscr{C}_{\mathbf{C}(x)} = \mathfrak{H} \cong \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}$ with generators*

$$\mathfrak{p} = (0, 0),$$
$$\mathfrak{q} = (x^2(x^2 - 3) + 2x, 0),$$
$$\mathfrak{s} = (x^2(x^2 - 1), 2x^2(x^2 - 1)).$$

In particular, $\mathscr{C}_{\mathbf{C}(x)} = \mathscr{C}_{\mathbf{Q}(x)}$.

Proof. Suppose first that $\mathscr{C}_{\mathbf{C}(x)}$ has rank 3. By Lemmas 5.3, 5.7, and 6.5, the field K is not real and there is a point $\mathfrak{c} \in \mathscr{C}_{\mathbf{C}(x)}$ defined over an imaginary quadratic field $\mathbf{Q}(\sqrt{(-d)}) \subset K$ with $\mathfrak{c} \to -\mathfrak{c}$ under the automorphism $\sqrt{(-d)} \to -\sqrt{(-d)}$. Here $d > 0$ is a square-free integer. The point $\mathfrak{c}$ corresponds to a point (ξ, η) on

$$\mathscr{C}^{-d} : -d\eta^2 = \xi(\xi - x^2(x^2 - 3) - 2x)(\xi - x^2(x^2 - 3) + 2x), \quad (7.1)$$

with $\xi, \eta \in \mathbf{Q}(x)$.

By Lemma 5.4 we may suppose, on replacing $\mathfrak{z}$ by $\mathfrak{z} - \mathfrak{h}$, where $\mathfrak{h}$ is a point of order 2, if need be, that $\xi \in \mathbf{R}(x)^{*2}$. (Note that $\mathfrak{h}$ is defined over $\mathbf{Q}(x)$ and so goes over into $\mathfrak{h} = -\mathfrak{h}$ under the automorphism.) This implies

$$\xi = c \cdot \frac{\theta^2}{\psi^2}, \quad (7.2)$$

where $\theta, \psi \in \mathbf{Q}[x]$, g.c.d. $(\theta, \psi) = 1$, and $c > 0$ is a square-free integer. Hence

$$c((c\theta^2 - x^2(x^2 - 3)\psi^2)^2 - 4x^2\psi^4) = -d\lambda^2 \quad (7.3)$$

for some $\lambda \in \mathbf{Q}[x]$.

Comparing constant terms in (7.3) shows that $\theta = x\theta_1$, $\lambda = x\lambda_1$, and $x \nmid \psi$. Then

$$c(x^2(c\theta_1^2 - (x^2 - 3)\psi^2)^2 - 4\psi^4) = -d\lambda_1^2. \quad (7.4)$$

On specializing x to 0, this implies $c = d$, and on looking at terms of highest degree we have

$$\deg(c\theta_1^2 - (x^2 - 3)\psi^2) < \deg \psi^2, \quad (7.5)$$

so

$$c = d = 1.$$

Thus (ξ, η) is a $\mathbf{Q}(x)$-rational point on $\mathscr{C}^{-1}$ with $\xi \in \mathbf{Q}(x)^{*2}$ in contradiction to Corollary 2.2 of Theorem 2.1.

Secondly, suppose that $\mathscr{C}_{\mathbf{C}(x)}$ has rank 2 and that K is not real. Then $K = Q(\sqrt{(-d)})$ with $d > 0$ and the proof proceeds as in the first case.

Suppose finally that $\mathscr{C}_{\mathbf{C}(x)}$ has rank 2 and that K is real. Then either $K = \mathbf{Q}(\sqrt{d})$ is real quadratic or $K = \mathbf{Q}$ (when we put $d = 1$). There is a point (ξ, η) on

$$\mathscr{C}^d : d\eta^2 = \xi(\xi^2 - 2x^2(x^2 - 3)\xi + x^2(x^2 - 1)^2(x^2 - 4)) \quad (7.6)$$

with $\xi, \eta \in \mathbf{Q}(x)^*$.

In addition, by Lemma 5.1 we may suppose, on adding an appropriate point of order 2 if necessary, that

$$\xi \in x(x-1)\,\mathbf{R}(x)^{*2} \quad \text{or} \quad \xi \in x(x+1)\,\mathbf{R}(x)^{*2}.$$

We suppose $\xi \in x(x-1)\,R(x)^{*2}$, on replacing x by $-x$ if need be. Then

$$\xi = cx(x-1)\frac{\theta^2}{\psi^2}, \tag{7.7}$$

where $\theta, \psi \in \mathbf{Q}[x]$, g.c.d. $(x(x-1)\,\theta, \psi) = 1$, and $c > 0$ is a square-free integer. From (7.6) we find

$$c(c^2x(x-1)\,\theta^4 - 2x^2(x^2-3)\,c\theta^2\psi^2 + x(x-1)(x+1)^2\,(x^2-4)\,\psi^4) = d\lambda^2 \tag{7.8}$$

for some $\lambda \in \mathbf{Q}[x]$.

Comparison of lowest terms in x shows that

$$c^2\theta(0)^4 - 4\psi(0)^4 = 0,$$

so

$$c = 2.$$

This implies that the highest terms in x on the left cannot cancel, so $d = 2$. On the other hand

$$4c^2\theta(1)^2\,\psi(1)^2 = d\lambda(1)^2.$$

This is possible only if $\theta(1) = \lambda(1) = 0$; and then the r.h.s of (7.8) is divisible by $(x-1)^2$, whereas the l.h.s is not: a contradiction.

Thus the rank of $\mathscr{C}_{\mathbf{C}(x)}$ must be 1 and then the result follows from Lemma 4.2.

8. Consequences for the Hasse Principle of Quadratic Forms and of Elliptic Curves

Our result that $f(x, y) = 1 + x^2(x^2-3)\,y^2 + x^2y^4$ is not a sum of 3 squares in $\mathbf{R}(x, y)$ is of interest for the general theory of quadratic forms in function fields. Similarly, our curve $\mathscr{C}$ serves as an example for the general behavior of elliptic curves over function fields. In both cases, one has the following notion of Hasse principle. Let K be an algebraic function field of transcendence degree 1 over a field k. Denote by $\mathfrak{p}$ the inequivalent valuations of K/k, by $K_\mathfrak{p}$ the completion of K with respect to $\mathfrak{p}$. We say that the Hasse principle for quadratic forms resp. elliptic curves holds in K if every quadric hypersurface resp. elliptic curve over K which has

points in all completions $K_\mathfrak{p}$ ("everywhere locally") has a point in K ("globally").

8.1. First, consider the case of quadratic forms and let $k = \mathbf{R}(x)$, $K = k(y)$. It can be shown (for details see [11]) that all completions $K_\mathfrak{p}$ of K have the property that any sum of squares in $K_\mathfrak{p}$ is a sum of three squares. In particular, our polynomial $f(x, y)$ is a sum of three squares in every $K_\mathfrak{p}$. But it is not a sum of three squares in K. Hence, the Hasse principle in K does not hold for the quadric

$$t_1^2 + t_2^2 + t_3^2 - ft_4^2 = 0.$$

In contrast to this the Hasse principle in K is true for

$$t_1^2 + \cdots + t_n^2 - at_{n+1}^2 = 0, \qquad a \in K^*$$

whenever $n \neq 3$. This is proved in Ref. [11] for $n = 1, 2, 4$ and follows trivially for $n > 4$, from the case $n = 4$. Another proof can be deduced from a recent result of G. Harder (unpublished) who shows that in every rational functional field $K = k(y)$, char $k \neq 2$, the following weaker form of Hasse's principle for quadratic forms holds: Two quadratic forms ϕ and ψ over K are equivalent over K if and only if they are equivalent over all $K_\mathfrak{p}$.

It has been known at least since Witt's paper [15] that the Hasse principle for quadratic forms is not generally true in function fields. This is rather trivial if the function field has genus $\geqslant 1$. Witt's example is a field of genus 0 over $k = \mathbf{Q}$. But it seems to be new that the Hasse principle fails already in such a "simple" rational function field as $\mathbf{R}(x, y)$.

8.2. We turn now to the Hasse principle for elliptic curves. Again let K/k be a function field in one variable and let $\mathscr{E}$ be an elliptic curve with rational point $\mathfrak{o}$ defined over K. We are interested in the Tate–Šafarevič group $Ш = Ш(\mathscr{E}, K)$ (for definition see, e.g., [3]) because Ш measures the validity of the Hasse principle for elliptic curves. More precisely, $Ш = 0$ if and only if the Hasse principle holds for all elliptic curves defined over K which become isomorphic to $\mathscr{E}$ over the separable algebraic closure of K.

In our case $K = \mathbf{R}(x)$ or $K = \mathbf{C}(x)$ and $\mathscr{E}$ is the curve (4.1). From our treatment in Section 7 we can easily deduce that the curve

$$\mathscr{E} : \eta^2 = x(x-1)\,\xi^4 - 2x^2(x^2-3)\,\xi^2 + x(x-1)(x+1)^2\,(x^2-4)$$

corresponds to a nonzero element (of order 2) of Ш. $\mathscr{E}$ has no K-rational point, since such a point would correspond to a K-rational point (ξ, η) on

$\mathscr{C}$, where ξ has square-class $x(x-1)$. On the other hand, we can show that $\mathscr{E}$ has points in every completion of K.

On clearing denominators the equation in question is

$$x(x-1)\,\theta^4 - 2x^2(x^2-3)\,\theta^2\psi^2 + x(x-1)(x+1)^2\,(x^2-4)\,\psi^4 = \lambda^2.$$

First, consider formal power series solutions in $x - x_0$ with $0 < x_0 < 1$. We get a solution by putting $\theta = 0$. If $x_0 = 0$ we can take $\theta = \sqrt{2}$, $\psi = 1$, if $x_0 = 1$, $\theta = \psi = 1$ will do. For the remaining places, that is the real places with $x_0 < 0$ or $x_0 > 1$, the infinite place and the complex places, one can put $\psi = 0$.

In conclusion it should be pointed out that there is an essential difference in the properties of Ш over number fields and over function fields. Over number fields Ш is conjectured to be finite whereas over function fields it may contain infinitely divisible elements (see [9], [13]). In fact, it follows from Theorems 3, 4, and 5 of Ref. [13] that Ш is infinitely divisible for our curve $\mathscr{C}$ and $K = \mathbf{C}(x)$. The result that $\mathscr{C}_K$ has rank 1 implies then that the 2-component of Ш is isomorphic to the direct sum of two copies of $\mathbf{Q}_2/\mathbf{Z}_2$.

References

1. J. Ax, On ternary definite rational functions (unpublished).
2. J. W. S. Cassels, On the representation of rational functions as sums of squares, *Acta Arith.* **9** (1964), 79–82.
3. J. W. S. Cassels, Diophantine equations with special reference to elliptic curves, *J. London Math. Soc.* **41** (1966), 193–291.
4. H. Davenport, A problematic identity, *Mathematika* **10** (1963), 10–12.
5. D. Hilbert, Über die Darstellung definiter Formen als Summe von Formenquadraten, *Math. Ann.* **32** (1888), 342–350 = Ges. Abh. **2**, 154–161.
6. D. Hilbert, Über ternäre definite Formen, *Acta Math.* **17** (1893), 169–197 = Ges. Abh. **2**, 345–366.
7. S. Lang, "Diophantine Geometry," Interscience, New York, 1962.
8. T. S. Motzkin, The arithmetic-geometric inequality, in "Inequalities," (Oved Shisha, ed.), pp. 205–224, Academic Press, New York, 1967.
9. A. P. Ogg, Cohomology of abelian varieties over function fields, *Ann. of Math.* **76** (1962), 185–212.
10. A. Pfister, Zur Darstellung definiter Funktionen als Summe von Quadraten, *Invent. Math.* **4** (1967), 229–237.
11. A. Pfister, Sums of squares in the function field $\mathbf{R}(x, y)$, in "The Proceedings of the Atlas Symposium No. 2," Academic Press, London (to appear).
12. R. M. Robinson, Some definite polynomials which are not sums of squares of real polynomials (To appear).
13. И. Р. Шафаревич, Главные однородные пространства определенные над полем функций, Труды Мат. Инст. им. Стеклова. **64** (1961), 316–346, (*Amer. Math. Soc. Transl.* **37** (1964), 85–114).

14. J. TATE, "Rational points on elliptic curves," Philips Lectures, Haverford College, 1961.
15. E. WITT, Über ein Gegenbeispiel zum Normensatz, *Math. Z.* **39** (1935), 462–467.
16. E. LANDAU, Über die Darstellung definiter Funktionen durch Quadrate, *Math. Ann.* **62** (1906), 272–285.